Community Dent Oral Epidemiol 2005; 33: 235–6

Community Dentistry and Oral Epidemiology

Letter to the Editor

In a recent manuscript, Lewsey and Thomson (1) investigated the application of zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression methods for modeling both cross-sectional and longitudinal dental caries data obtained from the same cohort study. Both ZIP and ZINB models incorporate extra variation than does the Poisson or negative binomial models. They have shown that the ZINB regression model provides a better fit compared with ZIP model for the dental caries data. After fitting some models to the data, it is essential to check the overall fit as well as quality of the fit. The quality of the fit between the observed values (y) and predicted values ($\hat{\mu}$) can be measured by various test statistics, however, the most two useful statistics are called scaled deviance (G^2 statistics) and Pearson deviance (χ^2 statistic) and are presented, respectively, below:

$$G^2 = \sum 2\left[y \log\left\{\frac{y}{\hat{\mu}}\right\} - (y - \hat{\mu})\right]$$

$$\chi^2 = \sum \frac{(y - \hat{\mu})^2}{\hat{\mu}}.$$

For a perfect fit model, both statistics G^2 and χ^2 have approximate chi-squared distribution with $(n-p)$ degrees of freedoms, where n is the number of observed data and p is the estimated parameters for the fitted model. For details about the fitting of a generalized linear model (GLM) readers are referred to McCullagh and Nelder (2) and more recently Wood (3) among others. As a goodness-of-fit of the models the authors have presented some graphs, which visualized the fit of the model. However, for a strong evidence and more confidence about the quality of the fit it is important to use goodness-of-fit test for this kind of analysis. The most appropriate reference for ZIP regression model is Lambert (4) and for ZINB regression model is Cameron and Trivedi (5) among others.

This paper has some minor errors. In the Appendix (p. 189), the ZIP model has been defined as

$$P(y|x) = \begin{cases} \pi + (1-\pi)\frac{e^{-\mu}\mu^y}{y!} & \text{for } y = 0 \\ (1-\pi)\frac{e^{-\mu}\mu^y}{y!} & \text{for } y > 0, \end{cases}$$

where y denotes the dependent variable, π is the probability of being an extra zero and $\mu = x\beta$, with x representing the independent variables and β the coefficients associated with x.

The correct model would be described as follows. The ZIP regression model can be expressed as

$$P(y|x) = \begin{cases} \pi + (1-\pi)e^{-\mu} & \text{for } y = 0 \\ (1-\pi)\frac{e^{-\mu}\mu^y}{y!} & \text{for } y > 0, \end{cases}$$

where y denotes the number of counts (or individuals) for a particular period or region, π $(0 < \pi < 1)$ is the probability of being an extra zero and μ is the expected number of occurrences per period, which can be modeled as

$$\mu = \exp(x'\beta),$$

where x' is the vector of explanatory variables and β is the vector of parameters. The remaining discussions of the Appendix are correct. This is a minor correction, which does not change the results or conclusion of the paper. The authors have used STATA (Stata 7.0; Stata Corporation, College Station, TX, USA) for analyzing the data.

References

1. Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. Community Dent Oral Epidemiol 2004;32:183–9.

2. McCullagh P, Nelder JA. Generalized linear models. London: Chapman and Hall; 1983.
3. Wood GR. Generalized linear accident models and goodness of fit testing. Accid Anal Prev 2002;34:417–27.
4. Lambert D. Zero-inflated Poisson regression, with an application to defects manufacturing. Technometrics 1992;34:1–14.
5. Cameron C, Trivedi P. Regression analysis of count data. New York: Cambridge University Press; 1998.

B. M. Golam Kibria
Department of Statistics
Florida International University
University Park
Miami, FL 33199, USA
E-mail: kibriag@fiu.edu